

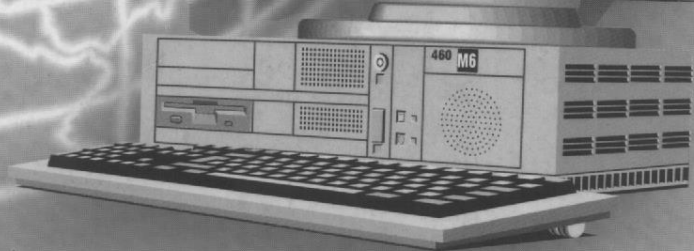
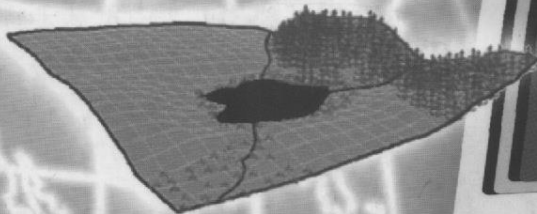
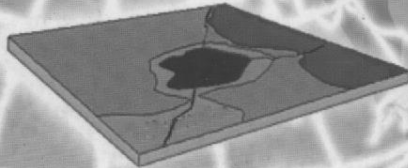
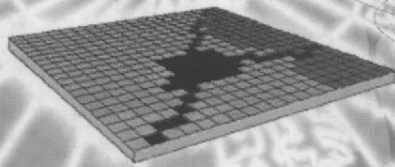


LIVRO  
TEXTO

# Sistemas de Informações Geo-referenciadas

Conceitos e fundamentos

Ardemirio de Barros Silva



  
EDITORA DA  
UNICAMP

## Capítulo 4

### CAPTURA DE DADOS PARA SIG

As variações geográficas do mundo real são infinitamente complexas. Quanto mais próximo está o observador mais detalhes podem ser vistos. É evidente que seria necessário um extraordinário e infinito banco de dados para conter precisamente os dados descritivos do mundo real. Dessa forma, os dados, de alguma maneira, precisam ser reduzidos em um número finito para poderem ser manipulados e essa redução se dá através de generalizações e abstrações. É sabido que curvas hipsométricas não existem no mundo real, entretanto são traçadas de forma semelhante lagos, casas e ruas, e estas feições correspondem a elementos do mundo real. Outros elementos são criados artificialmente, tais como os pixels. Também é importante salientar que algumas características do mundo real representam superfícies contínuas, isto é, cada valor encontrado na superfície da Terra depende, fundamentalmente, dos seus vizinhos, a exemplo da elevação, da temperatura, da declividade, da pressão atmosférica, entre outros.

A entrada ou captura de dados para alimentar o SGBD é uma das etapas do processo de análises de dados espaciais que requer cuidados especiais. A necessidade evidente de que os dados estejam em uma linguagem aceita pelos computadores, ou seja, em formato binário, torna essa tarefa difícil e exige procedimentos peculiares. Estes procedimentos, geralmente, correspondem a aproximadamente 75% dos esforços físicos, financeiros e intelectuais de um projeto envolvendo SIG. São quatro os tipos de dados que fazem parte de um SGBD: **textuais, numéricos, vetoriais e matriciais** ou **raster**.

O dado textual que tem caráter descritivo está em forma de texto e corresponde a fatos descritos sem qualquer interpretação. As informações numéricas correspondem a especificações codificadas em números do campo real, representando fatos do mundo real. Os dados textuais e numéricos podem ser eventualmente capturados utilizando-se do teclado do computador ou via *scanner*, produzindo um arquivo raster; em seguida, com o uso de um sistema de reconhecimento de caracteres (*optical character recognition* — OCR) o arquivo raster é transformado em formato em ASCII. A utilização do OCR traz várias vantagens: economia de tempo na geração dos arquivos ASCII (American

Standard Code for Information Interchange) e uma alta taxa de compressão dos dados. Entretanto o uso do OCR para capturar mapas é extremamente dificultado pela ocorrência de diferentes tipos e tamanhos de fontes, além de os objetos ocuparem posições as mais diversas. Os esquemas mais comuns de OCR são: *matrix matching*, que compara os arquivos raster gerados com uma biblioteca de arquivos de caracteres conhecidos; extração de atributos (*feature extraction*), que examina os arranjos de bits que compõem a letra.

Existem diferenças fundamentais entre os dados que descrevem o mundo real. O dado textual refere-se à descrição mais completa possível de construção mais concisa de fatos do mundo real, evitando-se ao máximo qualquer caráter interpretativo. O dado numérico corresponde ao atributo, categoria ou variável não espacial de determinado fato concernente ao meio físico.

O dado numérico pode ser classificado em **nominal**, **ordinal**, **intercalado** e **dividido**. A variável nominal corresponde ao atributo que, representando um item específico, não necessariamente obedece a uma determinada seqüência, como, por exemplo, tipos de solos ou de árvores de uma determinada região. Os dados ordinais apresentam-se como listas de classes discretas, com uma certa ordem inerente, como, por exemplo, as classes de rios (primeira, segunda e terceira ordem refletem os diferentes afluentes que contribuem para a formação do rio principal), a classe dos níveis de educação (primário, secundário, universitário, pós-graduação), a classe das eras geológicas (Arqueano, Fanerozóico, Cenozóico).

A variável intercalada representa um determinado intervalo natural e que tem um significado particular, ou seja, a diferença entre 10° C e 20° C é a mesma que entre 90° C e 100° C, porém a sua representatividade física é diferente. Os dados divididos têm as mesmas características que os dados intercalados, acrescentando-se que, por definição, eles têm uma origem inicial, que é o ponto zero, por exemplo, renda *per capita*, a fração do peso do material que fica retida em uma peneira com determinada malha, a precipitação anual das chuvas. A grande maioria dos dados que faz parte de um SIG é do tipo nominal ou ordinal.

Os detalhes das conexões entre os objetos espaciais, como, por exemplo, as informações sobre quais são as áreas limitadas por segmentos e suas

O dado vetorial é a representação gráfica do mundo real através de sistemas de coordenadas, dessa forma, a unidade fundamental do dado vetorial é o par de coordenadas  $x, y$ . O dado raster ou matricial refere-se à representação gráfica do mundo real através de pixels (*picture element*) ou células, com forma poligonal regular, geralmente quadradas, que são definidos pelas suas posições em relação às colunas e linhas de uma malha (Figura 35). Em termos mais gerais, podemos dizer que o dado raster representa o que ocorre em todos os lugares. O dado vetorial representa onde determinado fenômeno ocorre, ou seja, expressa a localização de todos os objetos.

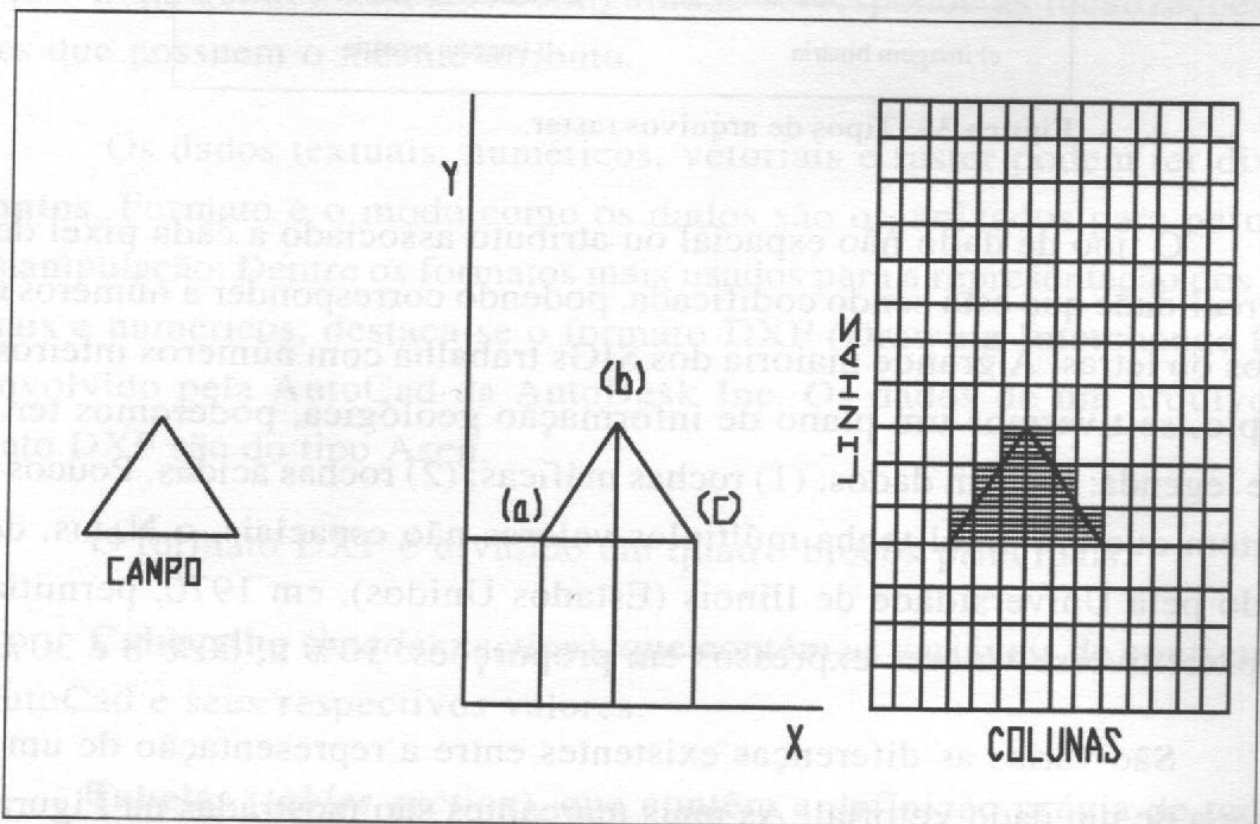
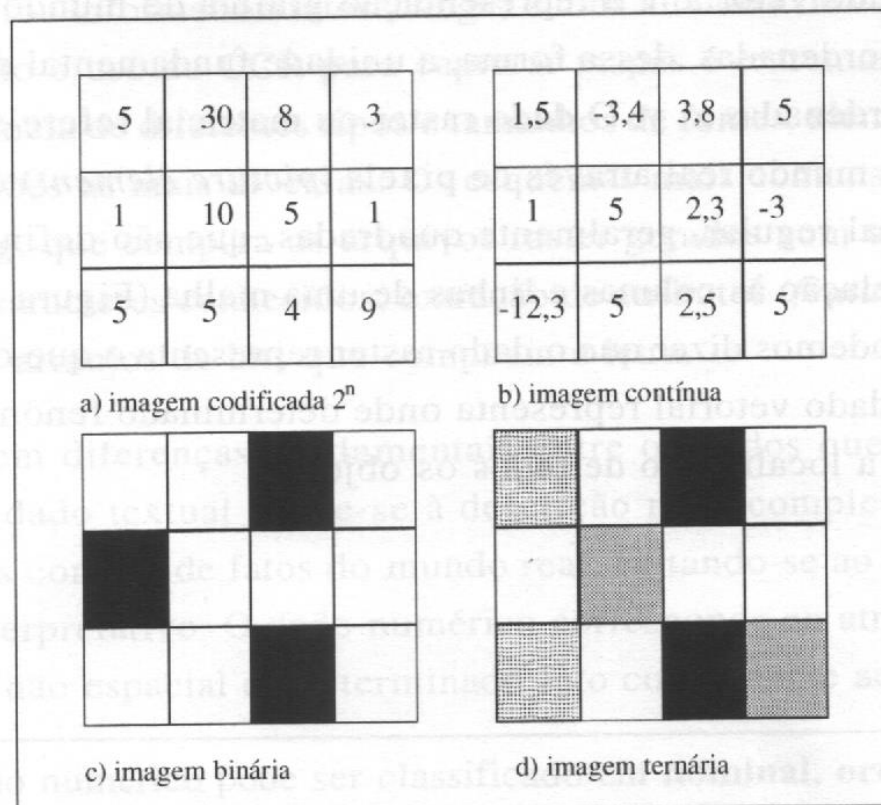


Figura 35: Representação de dado vetorial (1) e raster (2).

O dado codificado no formato raster pode ser classificado em: codificação  $2^n$ , imagem contínua, imagem binária e imagem ternária, tal como é mostrado na Figura 36.



**Figura 36:** Tipos de arquivos raster.

O tipo de dado não espacial ou atributo associado a cada pixel depende da realidade que está sendo codificada, podendo corresponder a números reais, inteiros ou letras. A grande maioria dos SIGs trabalha com números inteiros, por exemplo, se tivermos um plano de informação geológica, poderemos ter a seguinte legenda: (0) sem dados; (1) rochas máficas; (2) rochas ácidas. Poucos SIGs permitem que um pixel tenha múltiplos valores não espaciais, o NARIS, desenvolvido pela Universidade de Illinois (Estados Unidos), em 1970, permitia que cada pixel tivesse valores expressos em proporções: 10% a, 60% b e 30% c.

São várias as diferenças existentes entre a representação de um dado raster e a de um dado vetorial. As mais marcantes são mostradas na Figura 35 e referem-se a formas e medidas.

Com relação à forma, percebe-se claramente a distorção da forma triangular quando representada em formato raster. Em termos de medidas, se tomarmos como unidade de comprimento o lado do pixel, o triângulo representado vetorialmente apresenta um perímetro de  $6(1+2^{1/2})$  unidades e uma área de 9 unidades quadradas, enquanto o triângulo representado matricialmente apresenta um perímetro de 18 unidades e uma área 12 unidades quadradas, portanto a repre-

sentação raster, neste caso, apresentou um aumento de 20% no perímetro e 34% na área.

A codificação dos dados também é feita de maneira diversa. Enquanto a representação vetorial permite que as coordenadas  $x$ ,  $y$  possam ser expressas em números decimais, portanto muito mais detalhadamente, a representação raster sempre está codificada em números inteiros.

Os planos de informação (PI) possuem três características principais: resolução, orientação e contigüidade. A resolução é a mínima dimensão linear da menor unidade do espaço geográfico que foi registrada. A orientação corresponde ao ângulo entre o norte verdadeiro e a direção definida pelas colunas do modelo raster, no modelo vetorial ao ângulo entre o norte verdadeiro e qualquer linha contida no PI. A contigüidade corresponde às localizações adjacentes que possuem o mesmo atributo.

Os dados textuais, numéricos, vetoriais e raster podem ter diversos **formatos**. Formato é o modo como os dados são organizados para permitir a sua manipulação. Dentre os formatos mais usados para a representação dos dados textuais e numéricos, destaca-se o formato DXF (Drawing Interchange Files), desenvolvido pela AutoCad da AutoDesk Inc. Os dados de um arquivo com formato DXF são do tipo ASCII.

O formato DXF é dividido em quatro blocos principais:

**Cabeçalho** (*header section*), que contém as variáveis de configuração do AutoCad e seus respectivos valores.

**Tabelas** (*tables section*), que contém a definição prévia de todos os tipos de linhas, estilos de texto e níveis de informação (*layers*) a serem referenciados nas seções de blocos e entidades.

**Blocos** (*blocks section*), que contém a construção dos blocos que são entidades gráficas compostas por várias entidades primitivas utilizados no desenho.

**Entidades** (*entities section*), que correspondem às linhas, arcos, textos e inserção de blocos que armazenam as informações, via grupos de códigos (Tabela 12).

GRUPO DE CÓDIGOS	DESCRIÇÃO	GRUPO DE CÓDIGOS	DESCRIÇÃO
0	Identifica o início de uma entidade, tabela de entrada, ou separador de arquivo	21 - 28	Outras coordenadas Y
1	O valor do texto primário de uma entidade	30	Coordenada primária Z
2	Um nome; etiqueta de atributo, nome do bloco etc.	31 - 37	Outras coordenadas Z
3 - 4	Outros textos	38	Se não for zero a elevação da entidade
5	Entidade expressa como um texto hexadecimal	39	Se não for zero a espessura da entidade
6	Nome do tipo de linha	40 - 48	Valores de pontos decimais
7	Nome do estilo do texto	49	Valores repetidos
8	Nome dos níveis de informação ( <i>layer</i> )	50 - 58	Ângulos
9	Identificador do nome da variável	62	Número da cor
10	Coordenada primária do X (ponto inicial de uma linha, centro do círculo etc)	66	Outras entidades
11 - 18	Outras coordenadas X	70 - 78	Valores inteiros
20	Coordenada primária Y	210,220, 230	Componentes X, Y e Z de direção de extrusão

**Tabela 12:** Organização dos dados em formato DXF. O código do grupo 999 é destinado a comentários.

#### 4.1 Processos de captura de dados

É inerente a todo processo de captura de dados a existência de erros. Esses podem existir nos dados originais ou derivados da interpretação das fontes primárias. Entretanto é indispensável transformar todos os dados espaciais em formato digital, para que possam ser manipulados em ambiente de SIG. Em termos gerais, considera-se que a digitalização de um limite leva em torno de 1 min, assim 120 limites demandariam 2 horas e preços que variam entre R\$500,00 e R\$3000,00. Assim, é indispensável encontrar formas de redu-

zir os custos, aumentar a precisão. Uma das maneiras mais inteligentes é o compartilhamento entre várias empresas do dado digital.

Todos os processos utilizados na captura dos dados que se utilizam do teclado do computador para acionar comandos são conhecidos como **digitação**. O processo de captura de dados feito através de mouse e mesa digitalizadora é conhecido como **digitalização**. Existem diversos tipos de mouses no mercado e, genericamente, podem ser encontrados em dois tipos: com uma esfera rolante na parte inferior que se move quando o mouse é movido como um todo e outro com a esfera rolante na parte de cima acionada com o polegar. As vantagens entre um tipo e outro fica por conta dos operadores, não havendo uma razão técnica específica que aponte a escolha entre um ou outro.

Durante o processo de digitalização, o mouse é conectado a uma **mesa digitalizadora**, cujas dimensões podem variar desde 25 x 25 cm até 200 x 150 cm. Entre os vários tipos de mesas digitalizadoras, dois se destacam: uma mesa possuindo em sua base uma malha eletrônica ortogonal, e outra mesa do tipo ondas elétricas em fase. Para iniciar o processo de captura de dados, o mapa é fixado na mesa digitalizadora e pontos de controle são identificados, tais como interseção de estradas e pontos geodésicos. O número de pontos de controle depende da escala do mapa: quanto maior a escala, menor o número de pontos de controle a ser coletado.

Pontos são capturados acionando-se uma vez o *mouse*, retas, digitalizando-se o primeiro e o último ponto, curvas são capturadas deslocando-se lentamente o *mouse* do início ao fim da curva, e polígonos, partindo-se de uma origem, transferindo-se todas as retas até retornar ao ponto inicial.

Existem inúmeros problemas associados à captura de dados manual: o meio analógico é instável, cada vez que o mapa é removido da mesa, a nova sessão deve iniciar pelo reconhecimento dos pontos de controle; nem sempre os mapas revelam precisão, por exemplo, em vales estreitos, torna-se muito difícil a separação entre riachos, estradas de ferro e rodovias que ocorram simultaneamente no vale; a fadiga do operador deve ser levada em consideração; os mapas, com o passar do tempo, podem sofrer efeitos da dilatação térmica e, finalmente, nem sempre os mapas originalmente foram desenhados para serem digitalizados.



A todo processo de digitalização manual associam-se erros operacionais que devem ser quantificados, devendo estar dentro de intervalos admitidos como compatíveis com a escala do mapa utilizado. Durante o processo de captura de dados, numerosos erros podem ser encontrados, sendo o erro total o somatório de vários erros que pode ser expresso pela seguinte equação:

$$\text{ERRO TOTAL} = f(a) + f(m) + f(c) + f(d) + f(e) + f(p) + f(\text{RMS}) + f(\text{pe}) + f(\text{id}) + f(\text{at}) + f(\text{hi}) + I$$

Em que:

$f(a) \Rightarrow$  transformação da superfície esférica da Terra em superfície plana;

$f(m) \Rightarrow$  precisão da medida localizada na Terra (projeção cartográfica e *datum*);

$f(c) \Rightarrow$  interpretação cartográfica;

$f(d) \Rightarrow$  erro do desenho devido ao tamanho das linhas;

$f(e) \Rightarrow$  estabilidade do material (dobras, emendas etc.);

$f(p) \Rightarrow$  erro no processo de digitalização;

$f(\text{RMS}) \Rightarrow$  precisão do erro médio quadrático;

$f(\text{pe}) \Rightarrow$  precisão do equipamento (armazenamento de dados, número de dígitos);

$f(\text{id}) \Rightarrow$  idade do mapa;

$f(\text{at}) \Rightarrow$  descrição dos atributos ligados aos dados espaciais;

$f(\text{hi}) \Rightarrow$  dados históricos do mapa;

$I \Rightarrow$  erros adicionais não explicados.

Nos projetos desenvolvidos com SIG, é muito importante termos em mente que é particularmente interessante medir a representatividade dos dados. Entre os dados espaciais, sempre existem a incerteza e o erro. Na realidade, quando estamos realizando modelamentos e simulações, na maioria das vezes não se conhece os valores **exatos** dos dados, mas apenas valores considerados **precisos** ou **acurados**. Sem uma possibilidade de determinação do erro nessas circunstâncias, o termo que deve ser usado é a **incerteza**.

A incerteza é o conceito ou a condição de estar em dúvida sobre um valor. Percebe-se nesta definição o não-comprometimento com a **precisão**, a **acurácia** e a **exatidão**. A precisão é uma grandeza estatística que mede a dispersão, sendo, portanto, uma medida de repetibilidade das observações. A exatidão é a medida de aproximação entre um modelo matemático e o fenômeno representado. Uma distribuição possui acurácia se, dado algum intervalo de probabilidade, ela contém a resposta verdadeira. O erro, em seu sentido mais abrangente, é a diferença entre um valor medido ou calculado e o valor correto. A Figura 37 apresenta um gráfico mostrando as diferenças entre exatidão, precisão e acurácia.

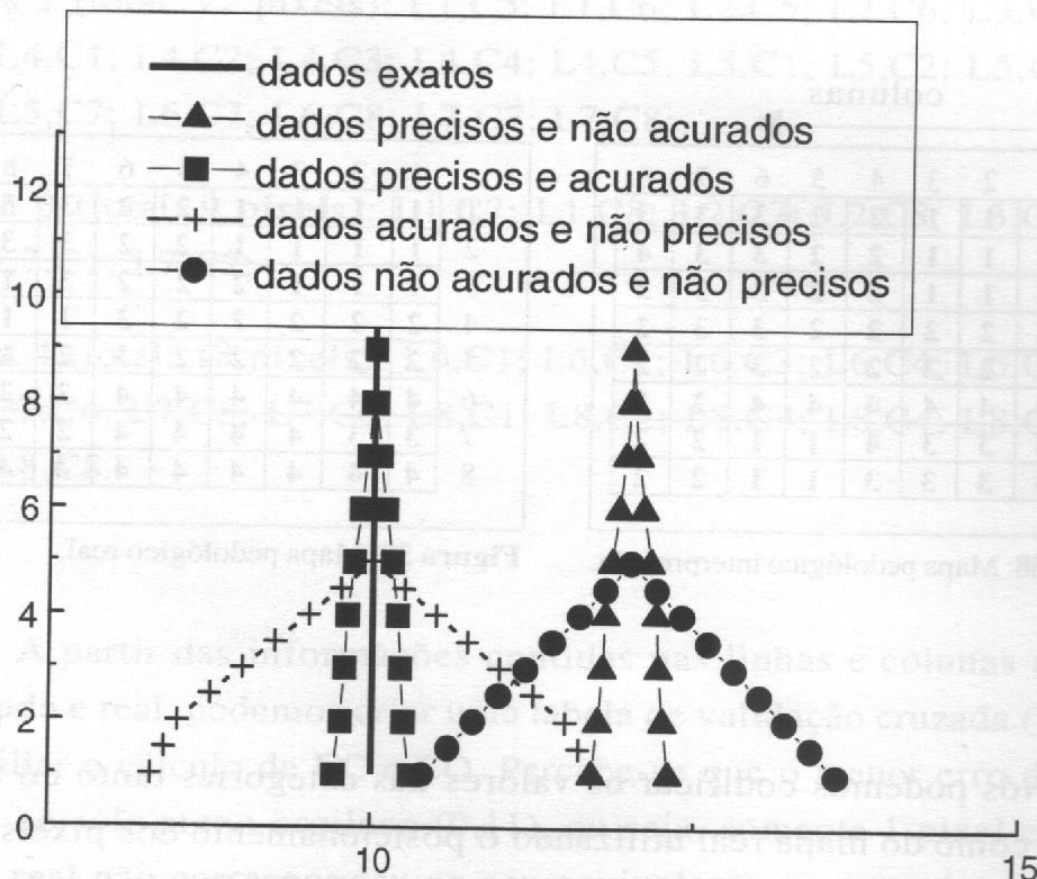


Figura 37: Representação esquemática da exatidão, precisão e acurácia.

Para garantir a qualidade dos dados em um SIG, é indispensável que sejam realizados testes para garantir a acurácia dos dados, uma vez que é praticamente impossível um SIG trabalhar com dados exatos. Um dos procedimentos mais simples é determinar os **erros de comissão (EC)** e **de omissão (EO)**, que variam entre 0 e 1 (maior erro). Erro de comissão é aquele decorrente da interpretação de pixels que não existem no terreno. Erro de omissão é a não-interpretação de pixels existentes no terreno.

Para exemplificar, imaginemos que um mapa de solo foi produzido a partir de interpretações de imagens de satélite, tendo sido identificados os seguintes tipos de solo: arenoso, argiloso, areno-argiloso e argilo-arenoso. O mapa pedológico produzido apresentou 64 pixels que representam solos arenoso (categoria 1), argiloso (categoria 2), areno-argiloso (categoria 3) e argilo-arenoso (categoria 4) tal como mostra a Figura 38. Foram realizadas etapas de campo e produzido um mapa pedológico (Figura 39) representando a realidade. Embora o número de pixels seja o mesmo, há diferenças entre os dois mapas. Para analisar quantitativamente a acurácia, foram calculados o EC e o EO.

colunas →

linhas ↓		1	2	3	4	5	6	7	8
	1	1	1	1	2	2	3	3	3
	2	1	1	1	2	2	3	3	4
	3	1	1	1	2	2	3	3	3
	4	2	2	2	2	2	3	3	2
	5	2	2	2	2	2	3	2	2
	6	4	4	4	4	4	4	2	2
	7	3	3	3	4	1	1	2	2
	8	3	3	3	3	1	1	2	2

Figura 38: Mapa pedológico interpretado.

	1	2	3	4	5	6	7	8
1	1	1	1	1	2	2	3	3
2	1	1	1	1	2	2	3	3
3	1	1	1	2	2	2	3	1
4	2	2	2	2	2	3	1	1
5	2	2	2	2	2	3	2	1
6	4	4	4	4	4	4	2	2
7	3	3	4	4	4	4	2	2
8	4	4	4	4	4	4	4	4

Figura 39: Mapa pedológico real.

Nós podemos codificar os valores das categorias tanto do mapa interpretado como do mapa real utilizando o posicionamento dos pixels em relação às linhas e às colunas.

Para o mapa interpretado, temos:

categoria 1 (total: 13 pixels): L1,C1; L1,C2; L1,C3; L2,C1; L2,C2; L2,C3; L3,C1; L3,C2; L3,C3; L7,C5; L7,C6; L8,C5; L8,C6;

categoria 2 (total: 25 pixels): L1,C4; L1,C5; L2,C4; L2,C5; L3,C4; L3,C5; L4,C1; L4,C2; L4,C3; L4,C4; L4,C5; L4,C8; L5,C1; L5,C2; L5,C3; L5,C4; L5,C5; L5,C7; L5,C8; L6,C7; L6,C8; L7,C7; L7,C8; L8,C7; L8,C8;

categoria 3 (total: 18 pixels): L1,C6; L1,C7; L1,C8; L2,C6; L2,C7; L3,C6; L3,C7; L3,C8; L4,C6; L4,C7; L5,C6; L7,C1; L7,C2; L7,C3; L8,C1; L8,C2; L8,C3; L8,C4;

categoria 4 (total: 8 pixels): L2,C8; L6,C1; L6,C2; L6,C3; L6,C4; L6,C5; L6,C6; L7,C4.

Para o mapa real, nós temos:

categoria 1 (total: 15 pixels): L1,C1; L1,C2; L1,C3; L1,C4; L2,C1; L2,C2; L2,C3; L2,C4; L3,C1; L3,C2; L3,C3; L3,C8; L4,C7; L4,C8; L5,C8;

categoria 2 (total: 22 pixels): L1,C5; L1,C6; L2,C5; L2,C6; L3,C4; L3,C5; L3,C6; L4,C1; L4,C2; L4,C3; L4,C4; L4,C5; L5,C1; L5,C2; L5,C3; L5,C4; L5,C5; L5,C7; L6,C7; L6,C8; L7,C7; L7,C8;

categoria 3 (total: 9 pixels): L1,C7; L1,C8; L2,C7; L2,C8; L3,C7; L4,C6; L5,C6; L7,C1; L7,C2;

categoria 4 (total: 18 pixels): L6,C1; L6,C2; L6,C3; L6,C4; L6,C5; L6,C6; L7,C3; L7,C4; L7,C5; L7,C6; L8,C1; L8,C2; L8,C3; L8,C4; L8,C5; L8,C6; L8,C7; L8,C8.

A partir das informações contidas nas linhas e colunas dos mapas interpretado e real, podemos criar uma tabela de validação cruzada (Tabela 12) para auxiliar o cálculo de EC e EO. Percebe-se que o menor erro de omissão refere-se ao solo areno-argiloso (0,11), ou seja, somente 1 pixel encontrado no mapa real não correspondeu ao seu equivalente no mapa interpretado. O

menor erro de comissão refere-se a solo argilo-arenoso (0,12), ou seja, somente 1 pixel interpretado não correspondeu ao pixel existente no campo.

	Arenoso (1)	Argiloso (2)	Areno-argiloso (3)	Argilo-arenoso (4)	TOTAL	ERRO DE COMISSÃO
Arenoso (1)	9	0	0	4	13	(1 - 9/13) 0,31
Argiloso (2)	4	19	0	2	25	(1 - 19/25) 0,24
Areno-argiloso (3)	2	3	8	5	18	(1 - 8/18) 0,55
Argilo-arenoso (4)	0	0	1	7	8	(1 - 7/8) 0,12
TOTAL	15	22	9	18	64	-
ERRO DE OMISSÃO	(1 - 9/15) <b>0,4</b>	(1 - 19/22) <b>0,14</b>	(1 - 8/9) <b>0,11</b>	(1 - 7/18) <b>0,61</b>	-	(1 - 43/64) <b>0,33</b>

**Tabela 13:** Utilizando os dados contidos nas Figuras 35 e 36, tabela de validação cruzada para cálculo de EC e EO.

Para avaliações mais elaboradas, recomendamos o cálculo do **Índice de Kappa** (Congalton e Mead, 1983), que mede a acurácia dos dados espaciais, desde que dados colecionados em campo sejam utilizados como referência. O Índice de Kappa (IK) varia entre 0 e 1, e os dados serão mais acurados quanto mais o índice se aproximar de 1.

Tomemos como exemplo duas matrizes 5x4, portanto com grau de liberdade igual a 19, uma representando um mapa de vegetação interpretado, seja a partir de imagens de satélite, seja produzido a partir de compilação bibliográfica (Figura 40), e outro mapa derivado de um exaustivo trabalho de campo (Figura 41). As categorias 1, 2, 3 e 4 representam mata nativa, reflorestamento, capoeira e cultura agrícola, respectivamente.

1	1	2	4	4
1	2	2	4	4
1	3	2	4	2
3	3	3	2	2

Figura 40: Mapa interpretado da vegetação.

1	2	2	2	4
1	3	2	2	2
3	3	3	2	2
3	3	3	3	2

Figura 41: Mapa de vegetação resultante dos trabalhos de campo.

Para o cálculo do IK, é necessário definir a correspondência existente entre os atributos dos pixels em suas respectivas posições espaciais. Assim, uma matriz de tabulação cruzada é construída com o objetivo de determinar as coincidências (Figura 42).

VERDADE DE CAMPO							
1	2	3	4		TOTAL		
2	1	1	0	1	4	BANCO	
0	4	3	0	2	7	DE	
0	0	4	0	3	4	DADOS	
0	4	0	1	4	5		
TOTAL	2	9	8	1	11		

Figura 42: Tabulação cruzada representando as coincidências entre o mapa de vegetação interpretado e o mapa derivado dos trabalhos de campo.

O total dos pixels examinados é 20 (T) e a soma dos valores encontrados na diagonal (D) representando a coincidência entre os pixels interpretados e os resultados dos trabalhos de campo é 11. O IK é definido pela seguinte equação:

$$IK = \frac{(D-Q)}{(T-Q)}$$

Onde D corresponde ao total de pixels com correspondência entre os trabalhos de campo e aqueles interpretados e T corresponde ao total de pixels examinados. Para o cálculo de Q, é utilizada a seguinte expressão:

$$Q = (2*4/20 + 9*7/20 + 8*4/20 + 1*5/20) = 5,4$$

Neste caso, o IK calculado é de 0,38, portanto os dados em exame não são acurados o suficiente para os modelamentos e simulações em SIG. Espera-se que dados possuindo um IK superior a 0,65 façam parte de banco de dados que serão utilizados pelos SIGs.

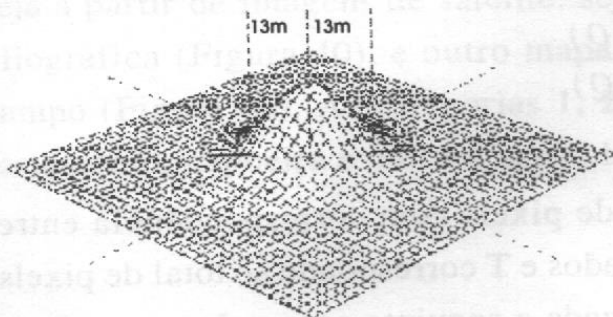
Para o cálculo do IK de uma única categoria, por exemplo, a categoria 2, pode-se fazer uso da seguinte expressão:

$$IK = (Q1 - Q2 * Q3) / (Q2 - Q2 * Q3),$$

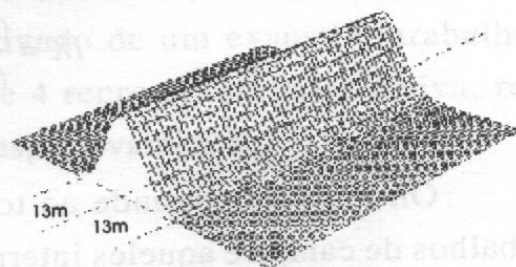
em que:  $Q1 = (4/20) = 0,2$ ;  $Q2 = (9/20) = 0,45$ ;  $Q3 = (7/20)$ ;

assim,  $IK = (0,2 - 0,45 * 0,35) / (0,45 - 0,45 * 0,35) = 0,14$ .

A quantificação do erro operacional é feita utilizando-se o erro médio quadrático (Root Mean Square — RMS) no eixo das ordenadas e das abscissas e, por fim, o cálculo do RMS total será representado graficamente como a hipotenusa de um triângulo retângulo. A Figura 43 apresenta um polígono capturado, a expressão gráfica do RMS e suas expressões matemáticas. O RMS permitido pode ser calculado segundo três perspectivas diferentes: a escala do mapa, a probabilidade de erro de um objetivo a ser alcançado e ao produto final esperado. Com relação à escala do mapa, em 1947, a United States National Map Accuracy Standards (USNMAS) (Padrões de Acurácia em Mapas Nacionais dos Estados Unidos) determinou que os mapas não deverão ter mais do que 10% de pontos testados, ou de verificação, com erros em faixas maiores que 1/30 polegadas em torno dos pontos verdadeiros, para mapas com escalas 1:20.000 ou menores (1:25.000), e em mapas com escalas maiores do que 1:20.000, por exemplo, 1:50.000, não deverão ter mais do que 10% de pontos testados com erros em faixas maiores que 1/50 polegadas em torno dos pontos verdadeiros, como pode ser visto nas Figuras 43 e 44.



**Figura 43:** Probabilidade de distribuição de um ponto. Para um mapa na escala 1:25.000, existem 90% de chance de o ponto se posicionar em um círculo de raio 13 m partindo do ponto real.



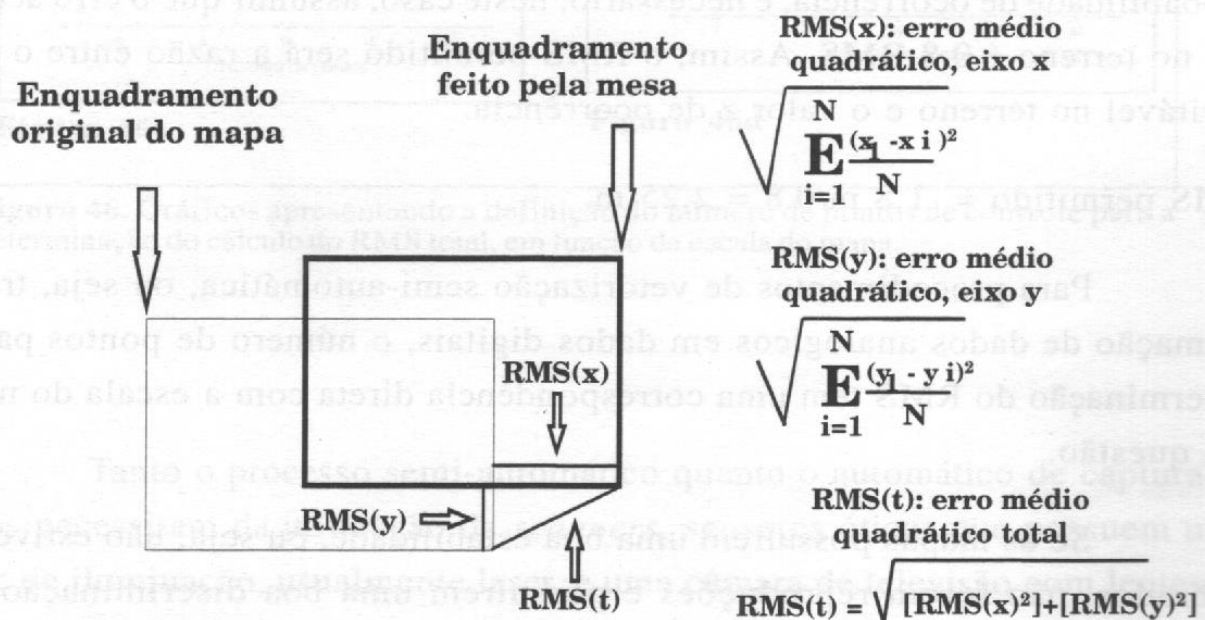
**Figura 44:** Probabilidade de distribuição de uma linha. Para um mapa na escala 1:25.000, existem 90% de chance de a linha se colocar em uma distância de até 13 m da linha central.

Atualmente, em virtude da formidável mudança nas técnicas de coleta de dados, é aqui sugerido que as faixas correspondam a 1/50 polegadas para mapas com escalas 1:20.000 ou menores e 1/60 polegadas para mapas com escalas maiores que 1:20.000. A conversão para padrões de acurácia é feita baseada em estatística. O RMS permitido requer que 90% de erros acidentais não devem ser maiores que 1,64% do RMS calculado, isto é, 1,64 desvios padrões, assumindo uma distribuição normal dos erros.

Por exemplo, se nós tivermos um mapa com escala de 1:15.000, o erro aceitável ou o RMS permitido será o erro aceitável multiplicado pela conversão da escala, multiplicado pela conversão da unidade, dividido por 1.64,

ou seja,  $\frac{1}{50} \text{ pol} * 15.000 * 0.0254 \text{ m/pol} = 4.64 \text{ m}$ . Para mapa com escala de 1:25.000,

o RMS permitido ou o erro aceitável será  $\frac{1}{60} \text{ pol} * 25.000 * 0.0254 \text{ m/pol} = 6.45 \text{ m}$



**Figura 45:** Representação do erro médio quadrático (RMS) nas abscissas, ordenadas e total.

Com relação ao RMS padrão, baseado na probabilidade de um objetivo a ser alcançado, pode ser descrito através de um exemplo. A prefeitura de uma cidade precisa registrar todas as posições das linhas telefônicas subterrâneas de tal forma que a probabilidade de que uma escavação, em um raio de 5 m em torno destas posições, possa interceptar, acidentalmente, uma li-



na não seja maior que 5 m. O valor  $z$  da probabilidade de ocorrência de 0,9999 é 3,9. Assim, o RMS ideal será calculado através da razão entre o valor do erro aceitável no terreno (5 m) e o valor  $z$  de probabilidade de ocorrência (3,9), ou seja,  $RMS = 1,3$ , portanto erros menores que 1,3 m asseguram a qualidade dos resultados.

Com relação ao RMS permitido, baseado em um produto final, por exemplo, um mapa. Imaginemos que um geólogo precisa definir os pontos que reflitam a localização de poços para extração de água subterrânea com um erro típico de 0,30 m (0,012 pol), e todos os pontos deverão estar plotados em um mapa na escala 1:5.000. Em um primeiro momento, devemos calcular o erro permitido no terreno que será calculado pela seguinte expressão:

Erro aceitável no terreno = Erro típico no mapa (0,30 mm) \* Escala(5000)\*Unidade de conversão(0,0012 m/pol) = 1,8.

Como o erro permitido depende do erro no terreno e do valor  $z$  de probabilidade de ocorrência, é necessário, neste caso, assumir que o erro aceitável no terreno é 0.8 RMS. Assim, o RMS permitido será a razão entre o erro aceitável no terreno e o valor  $z$  de ocorrência.

$RMS \text{ permitido} = 1,8 \text{ m} / 0,8 = 2,25 \text{ m}$

Para procedimentos de vetorização semi-automática, ou seja, transformação de dados analógicos em dados digitais, o número de pontos para a determinação do RMS tem uma correspondência direta com a escala do mapa em questão.

Se os mapas possuem uma boa estabilidade, ou seja, não estiverem dobrados, não forem reproduções e possuem uma boa discriminação das entidades a serem capturadas, o total de pontos pode ser determinado a partir do exame da relação existente entre número de pontos (eixo  $x$ ) e o valor do RMS (eixo  $y$ ). Para mapas possuindo escalas 1:10.000, 1:25.000, 1:50.000 e 1:100.000, o número ideal de pontos para calibração é 5, 7, 10 e 12, respectivamente, conforme é apresentado nas Figuras 46a, 46b, 46c e 46d.

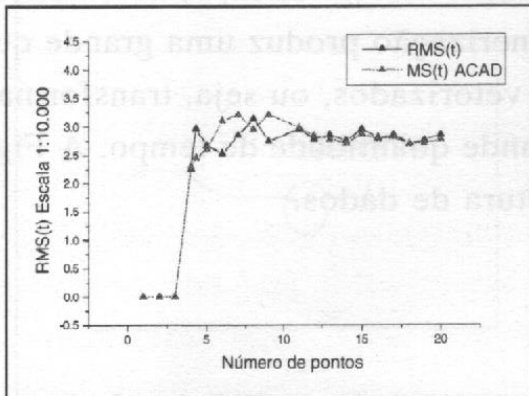


Figura 46a

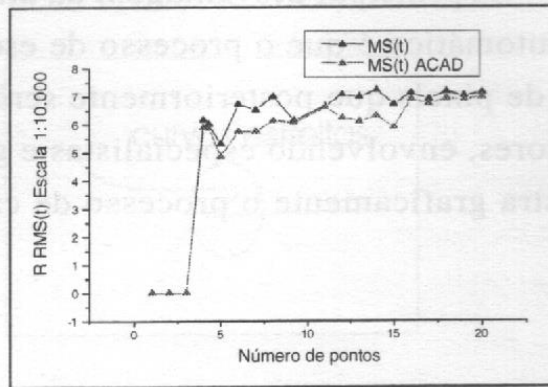


Figura 46b

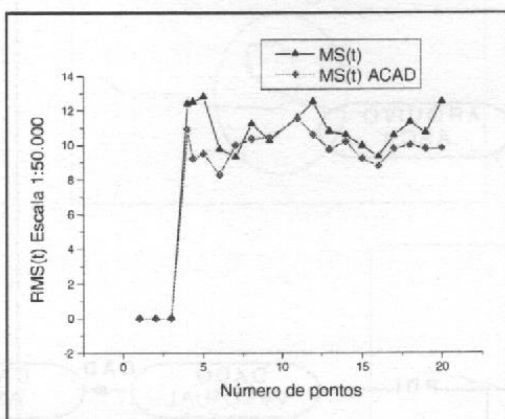


Figura 46c

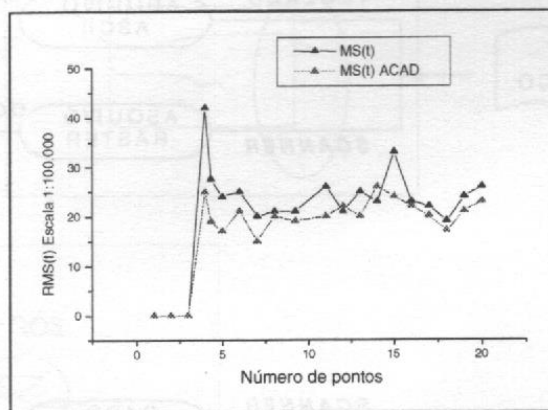
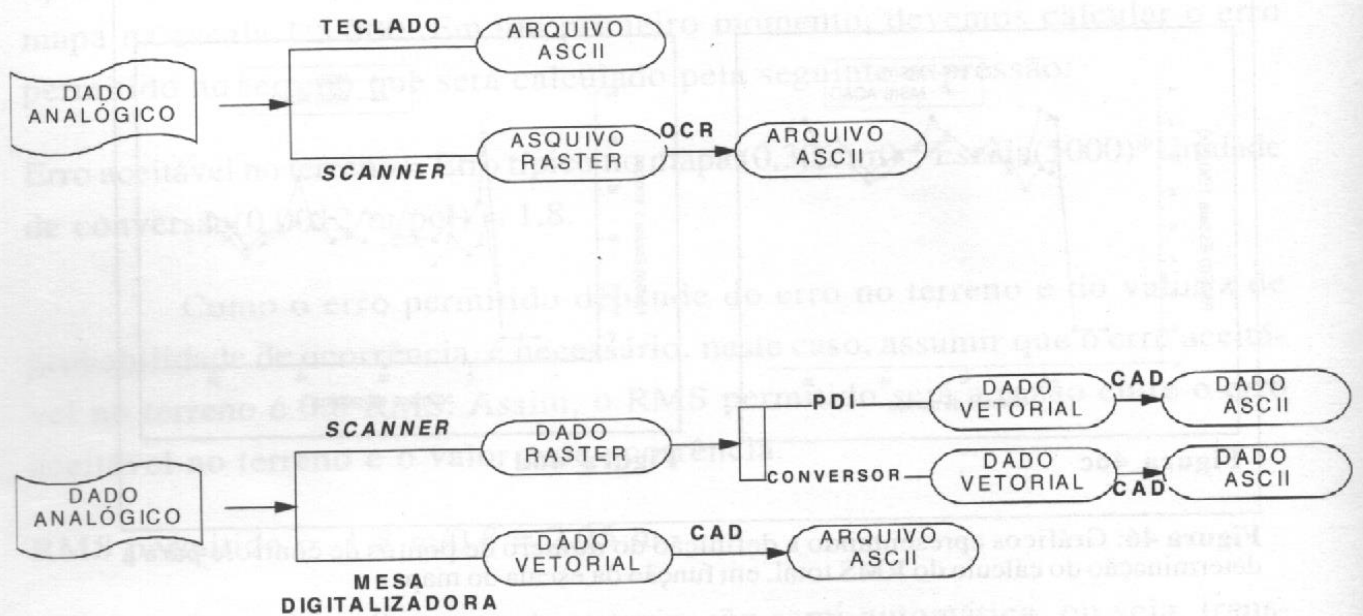


Figura 46d

**Figura 46:** Gráficos apresentando a definição do número de pontos de controle para a determinação do cálculo do RMS total, em função da escala do mapa.

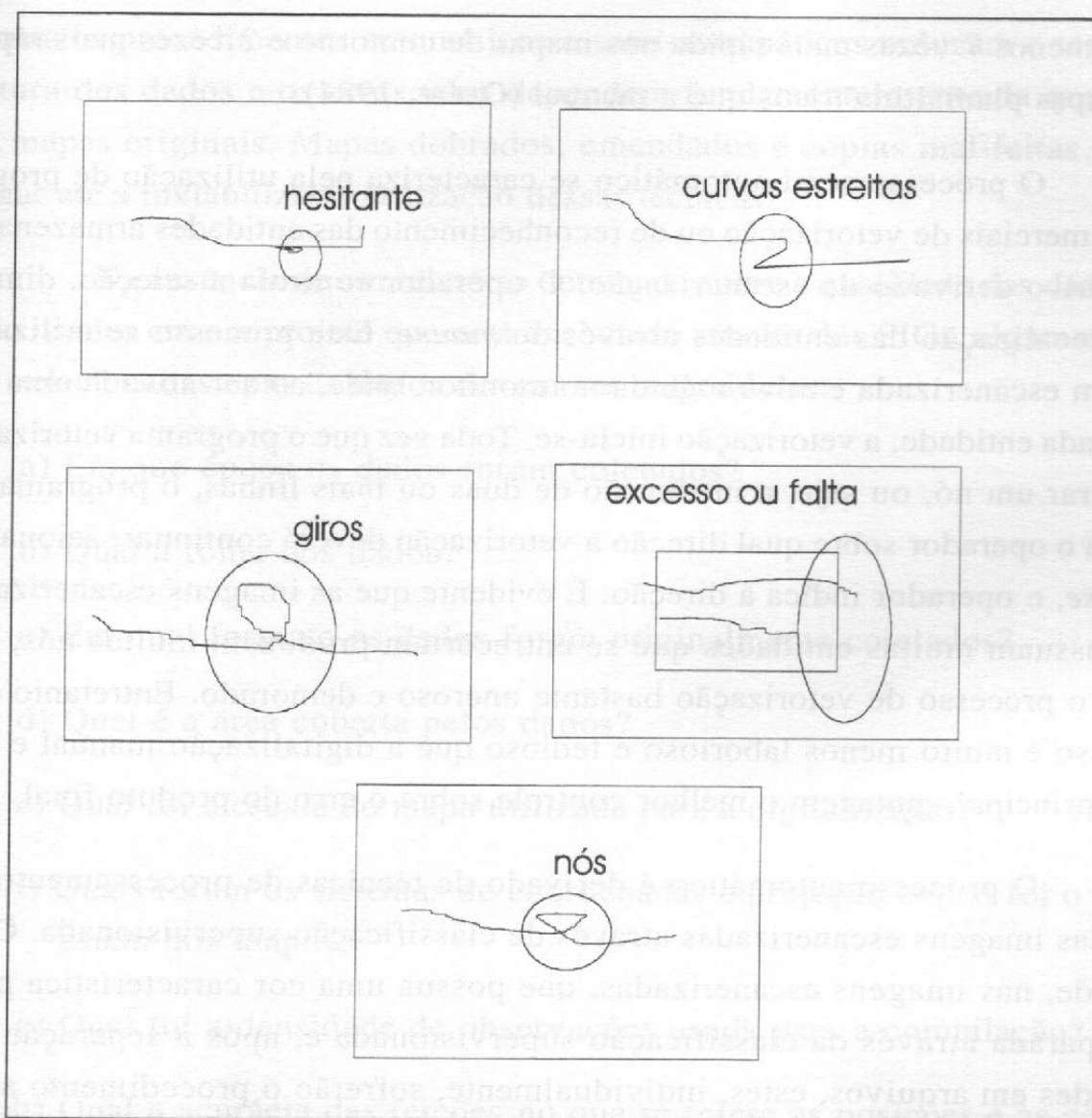
Tanto o processo semi-automático quanto o automático de captura de dados necessitam da utilização de *scanners*, sensores óticos que possuem uma fonte de iluminação, usualmente laser, e uma câmara de televisão com lentes de alta resolução que registram as informações contidas nos mapas analógicos em arquivos digitais com formato raster. Se o *scanner* fornece imagens com pixels de 25 micros (0,025 mm), uma linha com comprimento de 0,8 mm criará uma banda de 32 pixels. Os *scanners* podem ser manual, de mesa ou tambor, sendo este último um dos mais empregados, uma vez que minimiza as distorções, principalmente com relação às informações contidas nas bordas dos desenhos.

A principal desvantagem da utilização da digitalização automática ou semi-automática é que o processo de escanerização produz uma grande quantidade de pixels que posteriormente serão vetorizados, ou seja, transformados em vetores, envolvendo especialistas e grande quantidade de tempo. A Figura 47 ilustra graficamente o processo de captura de dados.



**Figura 47:** Procedimentos básicos utilizados na captura de dados analógicos.

Os chamados erros qualitativos são todos aqueles derivados de imprecisão do operador e podem ser classificados em hesitantes curvas estreitas, nós, excesso ou falta e giros, como são mostrados na Figura 48. É recomendável que, tendo finalizado todos os processos de vetorização, seja executada uma etapa de edição dos mapas para as correções dos chamados erros qualitativos, antes de migrá-los para um ambiente de SIG.



**Figura 48:** Principais erros qualitativos derivados dos processos de vetorização.

Para dar uma idéia do volume de dados gerados, tomemos como base um mapa de curvas hipsométricas, com uma dimensão de 30 x 50 cm, em uma região montanhosa; se escanizado com uma densidade de 1600 dpi (*dots per inch*), corresponderá a um arquivo com  $92 \times 10^6$  bits de dados, que, sem dúvida, além de ocupar um grande espaço na memória do computador, demandará muito tempo de processamento. Em contrapartida, uma grande vantagem é a redução do tempo de vetorização. Por exemplo, um mapa topográfico, na escala de 1:50.000, levará cerca de 30 minutos até 8 horas utilizando-se os procedimentos semi-automáticos e 4 a 6 dias utilizando os procedimentos manuais. Testes envolvendo vetorização automática e manual revelam que a automática

é pelo menos 7 vezes mais rápida nos mapas de contorno e 2 vezes mais rápida nos mapas planialtimétricos que a manual (OEEPE, 1984).

O processo semi-automático se caracteriza pela utilização de programas comerciais de vetorização ou de reconhecimento das entidades armazenadas no arquivo derivado da escanerização. O operador controla a seleção, dimensão e localização das entidades através do *mouse*. Este processo se utiliza da imagem escanerizada e ativa exibida no monitor, onde, ao ser ativada uma determinada entidade, a vetorização inicia-se. Toda vez que o programa vetorizador encontrar um nó, ou seja, a interseção de duas ou mais linhas, o programa interroga o operador sobre qual direção a vetorização deverá continuar; acionando o *mouse*, o operador indica a direção. É evidente que as imagens escanerizadas que possuam muitas entidades que se entrecortam produzem muitos nós, tornando o processo de vetorização bastante oneroso e demorado. Entretanto este processo é muito menos laborioso e tedioso que a digitalização manual e tem como principal vantagem o melhor controle sobre o erro do produto final.

O processo automático é derivado de técnicas de processamento digital das imagens escanerizadas através de classificação supervisionada. Cada entidade, nas imagens escanerizadas, que possua uma cor característica pode ser separada através da classificação supervisionada e, após a separação das entidades em arquivos, estes, individualmente, sofrerão o procedimento automático de vetorização. Os dados originais devem ser coloridos e o *scanner* de tambor capaz de produzir imagens também coloridas.

Alguns requisitos são necessários para escanerização de mapas. Mesmo utilizando-se do melhor *scanner* disponível e do mais atualizado programa vetorizador, a imagem resultante e o produto final podem conter imperfeições derivadas de sujeiras e defeitos contidos no mapa original. As linhas contidas nos mapas devem ter no mínimo 0,1 mm de largura. Linhas que estejam muito próximas podem ser codificadas como linhas grossas, ao invés de várias linhas finas. Linhas que possuam forma complexa têm muita chance de serem simplificadas no processo de escanerização.

Alguns dados espúrios podem aparecer no produto final escanerizado, os textos podem, acidentalmente, ser considerados linhas, símbolos especiais podem introduzir novos polígonos e linhas. É importante também destacar que,

embora os procedimentos automáticos e semi-automáticos reduzam o tempo de captura dos dados e os erros, eles dependem fundamentalmente da qualidade dos mapas originais. Mapas dobrados, emendados e cópias mal-feitas podem chegar até a inviabilizar a utilização dessas técnicas.

Na ausência de um relatório detalhado sobre a qualidade dos dados que serão usados em um projeto que será executado através dos SIGs, algumas questões relacionadas com os dados devem ser respondidas:

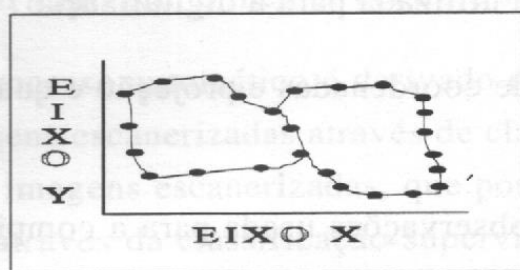
- a) Em que época os dados foram coletados?
- b) Qual a fonte dos dados?
- c) Em qual formato os dados foram originalmente coletados?
- d) Qual é a área coberta pelos dados?
- e) Qual foi a escala do mapa utilizada para a digitalização?
- f) Quais foram os sistemas de coordenadas e projeção e qual foi o *datum* usado nos mapas?
- g) Qual foi a densidade de observações usada para a compilação?
- h) Qual a acurácia das feições no que se refere às posições e às características dos atributos?
- i) Qual é a consistência e a lógica dos atributos?
- j) Como os dados foram testados?

## 4.2 Dados vetoriais

São várias as formas como os dados vetoriais podem ser estruturados e classificados, de acordo com a sua natureza. Assim temos as entidades **pontuais**, **lineares** e **poligonais**. As entidades pontuais podem ser representadas por um único par de coordenadas. As entidades lineares podem ser represen-

tadas por dois ou mais pares de coordenadas. As entidades poligonais correspondem a um conjunto de entidades lineares, cujos pares de coordenadas inicial e final são os mesmos. As entidades poligonais podem ser estruturadas em diferentes formatos. Os mais comuns são: **modelo total**, **modelo topológico**, **modelo espagete**, **Dual Independent Map Encoding (DIME)**, **modelo relacional** e **Digital Line Graphs (DGL)** (Star e Estes, J. E., 1990; Aronoff, 1989; Burrough, 1986).

O modelo total é a forma mais simples de estruturação. Cada polígono é codificado em coordenadas **X** e **Y** (Figura 49), não mantendo as relações matemáticas entre os objetos. Não há mecanismos explícitos para referenciar polígonos que sejam adjacentes. Os arquivos gerados em programas do tipo CADD pertencem a esta categoria.

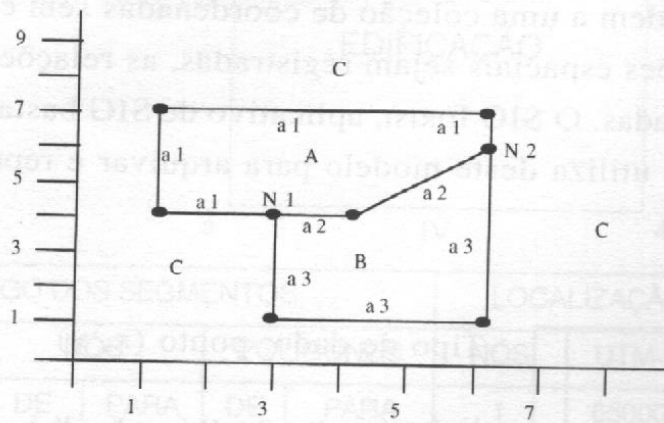


FEIÇÕES	NÚMERO	REPRESENTAÇÃO
POLÍGONOS	1	X1Y1... XnYn
	2	X1Y1... XnYn

**Figura 49:** Modelo total.

O modelo topológico (Figura 50) apresenta uma estrutura compacta. Seus elementos são divididos em arcos/linhas. O início, o término ou o encontro de arcos são denominados nós, as relações matemáticas entre os objetos são mantidas. Constitui-se um dos modelos mais comuns de representação de dados vetoriais. As relações matemáticas entre os objetos são registradas em tabelas: uma para áreas, outra para arcos e outra para nós. Em uma quarta ta-

bela, encontram-se registradas as coordenadas. Os arcos que definem os limites de cada polígono são codificados, por convenção, no sentido horário.



POLÍGONO	ARCOS
A	A1, a2
B	A2, a3

NÓ	ARCO
N1	a1, a2
N2	a1, a3

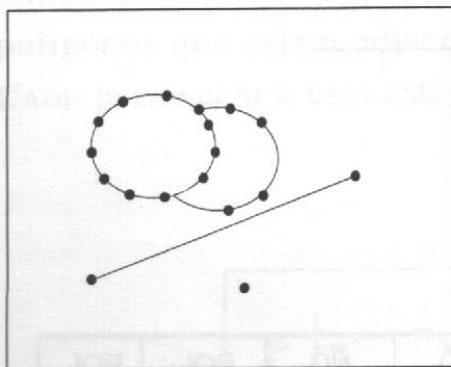
ARCO	NÓ INICIAL	NÓ FINAL	POL. ESQ.	POL. DIR.
a1	N1	N2	C	C
a2	N2	N1	A	B
a3	N2	N1	C	C

COORDENADAS DOS ARCOS			
ARCO	INÍCIO x, y	PONTO INTERMEDIÁRIO	FINAL x, y
a1	3,4	2,3; 2,5; 5,5; 5,4	6,7
a2	3,4	4,3; 5,4	5,4
a3	3,4	5,1; 3,1; 3,3	5,4

Figura 50: Modelo topológico.



O modelo espaguete (Figura 51), embora de fácil entendimento e boa adequação para a representação de mapas, apresenta dados redundantes, uma vez que as linhas comuns a dois polígonos são armazenadas duas vezes e as relações espaciais não são arquivadas. Essencialmente, os dados espaciais coletados correspondem a uma coleção de coordenadas sem estruturas inerentes. Embora as feições espaciais sejam registradas, as relações espaciais entre elas não são codificadas. O SIG IDRISI, aplicativo de SIG bastante difundido no meio acadêmico, se utiliza deste modelo para arquivar e representar os PIs.



Tipo de dado: ponto  $(x, y)$

linha  $(x_1, y_1; x_2, y_2; \dots x_n, y_n)$

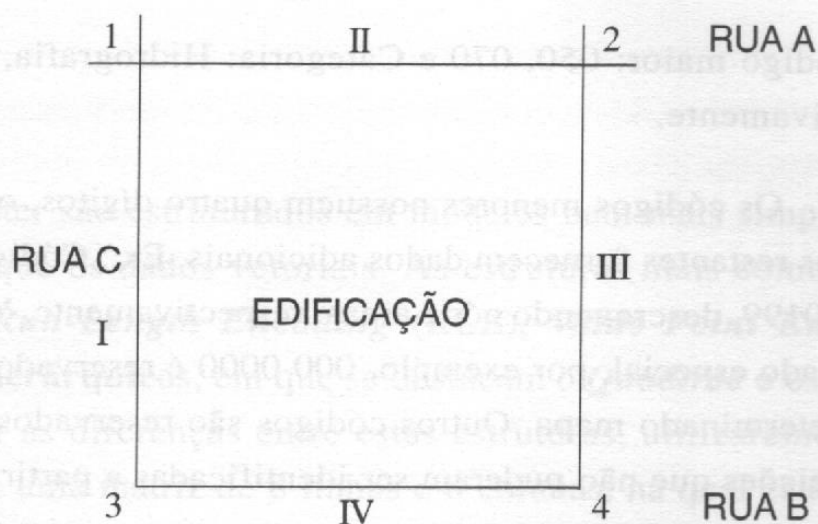
polígonos  $(x_1, y_1; x_2, y_2; \dots x_n, y_n)$

$(x_1, y_1; x_2, y_2; \dots x_n, y_n; w_1, z_1; w_2, z_2; \dots w_n, z_n)$

**Figura 51:** Representação de dados vetoriais segundo o modelo espaguete.

O DIME foi desenvolvido pelo US Bureau of the Census dos Estados Unidos, em 1967, e foi concebido para incorporar informações topológicas de áreas urbanas para uso em análises demográficas. Embora os arquivos tipo DIME geralmente não correspondam à organização interna de um SIG, eles são comumente usados como banco de dados que podem servir a diversos sistemas.

A estrutura dos dados é mostrada na Figura 52, no qual os segmentos lineares e os nós são compartilhados pelas unidades poligonais adjacentes. Quando são necessárias linhas curvas, elas são representadas por uma seqüência de segmentos retos. Cada segmento é armazenado possuindo três componentes essenciais: um nome para o segmento (ex.: o nome da rua) que o identifica, um nome identificando os nós, codificando o início e final dos segmentos e, finalmente, um outro componente indicativo dos polígonos, mostrando se eles estão do lado esquerdo ou direito de um determinado segmento.



CÓDIGO DOS SEGMENTOS					LOCALIZAÇÃO DOS NÓS		
NOME DO SEGMENTO	NÓS		POLÍGONOS		NÓS	UTM-E	UTM-W
	DE	PARA	DE	PARA	1		
RUA A	1	2	-	CAMPO	2	700000	7560000
RUA B	3	4	-	CAMPO	3	700000	7590000
RUA C	4	1	-	CAMPO	4	650000	7590000

Figura 52: Exemplo de modelo de estrutura de dados DIME.

O modelo DLG foi desenvolvido pela US Geological Survey. Os dados existentes na estruturação DLG são subdivididos em diversos níveis temáticos:

1. consistem de informações de limites que incluem os limites políticos e administrativos de uma região;
2. compreendem a rede de drenagem;
3. consistem na rede de transporte;
4. correspondem ao uso e ocupação das terras.

Os elementos essenciais como nós e arcos têm as mesmas definições já comentadas. Acrescenta-se a esses elementos essenciais a chamada linha degenerada que corresponde a uma linha de comprimento 0 (zero), usada para definir feições, que são indicadas no mapa como pontos.

Cada área pode ter um ponto associado que representa as características da área. Os códigos e atributos são estruturados de uma maneira específica. O código maior contém três dígitos, os primeiros dois dígitos representam a categoria geral do elemento e o terceiro fornece detalhes adicionais.

Ex.: Código maior: 050, 070 e Categoria: Hidrografia, Cobertura superficial, respectivamente.

Os códigos menores possuem quatro dígitos, o primeiro é geralmente zero e os restantes fornecem dados adicionais. Ex.: Código menor: 0001 – 0099, 0100 – 0199, descrevendo nós e áreas, respectivamente. Vários códigos possuem significado especial, por exemplo, 000 0000 é reservado para uma área externa a um determinado mapa. Outros códigos são reservados para informações, tais como feições que não puderam ser identificadas a partir dos dados disponíveis.

Em 1990, foi feita uma fusão do DIME e DLG para a geração de arquivos que foram usados nas análises do censo 90. A nova estruturação chamou-se TIGER (Topologically Integrated Geographic Encoding and Refering).

Existem vantagens e desvantagens na utilização da representação da estrutura dos dados vetoriais segundo este ou aquele modelo, de modo a atender a diferentes expectativas. A Tabela 14 apresenta, de maneira sintética, as vantagens e desvantagens da utilização das diversas estruturas de dados vetoriais.

TIPO	VANTAGENS	DESVANTAGENS
MODELO TOTAL	Estrutura compacta	Não há mecanismos explícitos para referenciar áreas adjacentes As relações matemáticas entre objetos não são arquivadas
MODELO TOPOLÓGICO	Estrutura compacta As relações matemáticas entre objetos são arquivadas Rápido processamento de dados	Estruturas complexas Redundância de dados
MODELO ESPAGUETE	Estrutura compacta e de fácil entendimento Adequado para mapas	Lento processamento de dados As relações espaciais não são arquivadas As linhas comuns a dois polígonos são arquivadas duas vezes
DIME	As estruturas e os atributos são arquivados separadamente	Redundância de dados Complexo gerenciamento de arquivos
DLG	São topologicamente estruturados Custo baixo	Estruturas complexas

**Tabela 14:** Vantagens e desvantagens da utilização das diversas estruturas de dados vetoriais.

### 4.3 Dados raster

Os dados raster são estruturados em modelos bem mais simples e de forma mais compacta que os dados vetoriais. As estruturas mais comuns são: **Código de Cadeias**, **Run-Length Encoding (RLE)**, **Value Point Encoding (VPE)** e os **Modelos Hierárquicos**, em que se destacam o **Quadtree** e os **Blocos Máximos**. Para ilustrar as diferenças entre estas estruturas, utilizaremos a Figura 53, que representa uma matriz de 8 linhas e 8 colunas, na qual estão codificadas três feições do mundo real, representadas pelas categorias 1, 2 e 3, que podem ser armazenadas em um arquivo contendo 64 unidades de bytes.

1	1	1	1	1	2	2	2
1	1	1	1	1	2	2	2
1	1	1	2	2	2	2	2
1	1	1	2	2	2	2	2
1	1	1	2	2	2	2	2
2	2	2	2	2	3	3	2
2	2	2	2	2	3	3	2
2	2	2	2	2	2	2	2

**Figura 53:** Representação matricial de 3 determinados tipos de solo (solos 1, 2 e 3).

O Código de Cadeias usa uma notação particular, os algarismos 0, 1, 2 e 3 servem como bases representando direções este, sul, oeste e norte, respectivamente. A partir da linha 1 e coluna 1, no sentido anti-horário, os dados são organizados segundo determinada categoria. Assim, parte da categoria 1 está codificada como: 1<sup>5</sup>, 0<sup>3</sup>, 3<sup>5</sup>, 2<sup>3</sup>. A estruturação em código de cadeias reduziu o tamanho do arquivo em 8 unidades de bytes.

A Tabela 15 apresenta a estruturação usando o Código de Cadeias, neste caso, o total de informações registradas corresponde a 56 unidades de bytes.

LINHA	COLUNA	ATRIBUTO	CODIFICAÇÃO
1	1	1	$1^5, 0^3, 3^5, 2^3$
1	4	1	$1^1, 0^1, 3^1, 2^1$
1	6	2	$1^5, 0^3, 3^5, 2^3$
3	4	2	$1^6, 0^2, 3^6, 2^2$
6	1	2	$1^3, 0^3, 3^3, 2^3$
6	6	3	$1^1, 0^1, 3^1, 2^1$
8	6	2	$1^0, 0^2, 3^0, 2^0$
8	8	2	$1^3, 0^0, 3^0, 2^0$

**Tabela 15:** Codificação em estrutura Código de Cadeias, a partir da Figura 53.

O princípio do RLE consiste em que pixels adjacentes tendo o mesmo atributo são combinados juntos, como uma tira, representada por um par de números. Cada nova linha inicia uma nova tira. O número de bits usados para o RLE é dependente do número de colunas. Assim, uma imagem com 1024 colunas precisa de 10 bits para as tiras, que podem chegar até o comprimento de  $2^{10}$  bytes. Tomando como imagem a Figura 53, a codificação em RLE seria: linha 1: (5, 1), (3, 2); linha 2: (5, 1), (3, 2); linha 3: (3, 1), (5, 2); linha 4: (3, 1), (5, 2); linha 5: (3, 1), (5, 2); linha 6: (5, 2), (2, 3), (1, 2); linha 7: (5, 2), (2, 3), (1, 2); linha 8: (8, 2). A codificação em RLE diminui sensivelmente o tamanho dos arquivos. Alguns SIGs, como o GRASS (Geographic Resources Analysis Support Systems), fazem uso desse benefício. É importante destacar que o sistema operacional UNIX se utiliza do RLE para compactar os arquivos.

O VPE codifica os dados a partir da linha 1 e coluna 1, continuamente, associando o atributo ao número de pixels. A cada mudança de atributo tem-se um determinado número de pixels. Utilizando-se da matriz apresentada na Figura 53, a codificação em VPE é representada da seguinte maneira: coluna 1, linha 1 5(1); 2(8); 1(13); 2(16); 1(19); 2(24); 1(27); 2(32); 1(35); 2(45); 3(47); 2(53); 3(55); 2(64).

O princípio fundamental do *quadtree* é a divisão do dado raster a partir de múltiplos polígonos regulares (quadrados). Assim, o primeiro quadrado, que compreende toda a matriz, contém 3 categorias, como mostra a Figura 54. Este é dividido em sucessivos quadrados, até a unidade fundamental, que é o pixel. O conceito que determina o número de quadrados é que em cada quadrado deve ser encontrada sempre a mesma categoria. A Figura 54 apresenta, de maneira esquemática, uma representação em *quadtree*, na qual pode-se perceber a divisão em diversos polígonos regulares (quadrados) de diferentes tamanhos. Assim, o maior quadrado envolve 4 pixels possuindo as categorias 1 e 2. A categoria 3 está presente em um *quadtree*, possuindo somente um pixel.

1	1	1	1	1	2	2	2
1	1	1	1	1	2	2	2
1	1	1	2	2	2	2	2
1	1	1	2	2	2	2	2
1	1	1	2	2	2	2	2
2	2	2	2	2	3	3	2
2	2	2	2	2	3	3	2
2	2	2	2	2	2	2	2

**Figura 54:** Representação gráfica da codificação de dado raster em estrutura *quadtree*.

O modelo de Blocos Máximos é uma variante do *quadtree*, cujo objetivo é eliminar possíveis redundâncias. A matriz é subdividida em polígonos regulares, não necessariamente quadrados, e cada polígono contém um único atributo.

Os métodos de estruturação de dados raster apresentam diferentes vantagens e desvantagens, tal como é mostrado na Tabela 16.

MÉTODOS DE COMPRESSÃO	VANTAGENS	DESVANTAGENS
RLE E VPE	São voltados para pequenos computadores Estruturas simples	Menos eficiente quando o limite entre os polígonos aumenta Aumenta os tamanhos dos arquivos e o tempo de processamento quando o número de polígonos aumenta
CÓDIGO DE CADEIAS	Estruturas simples Dados compactados	Presença de redundância; os limites dos polígonos são armazenados duas vezes Superposições são difíceis sem que se retorne à grade original
BLOCOS MÁXIMOS	Eficiente para formas simples e grandes	Limitado para ser usado em arquivos com muita variabilidade espacial
QUADTREE	Estrutura elegante Dados compactados Rápido processamento	Dificuldade para criação e atualização de arquivos

**Tabela 16:** Vantagens e desvantagens do uso das estruturas de dados raster.

Exemplos típicos de dados raster são matriz de elevação, classes de uso de solos, dados geológicos, geofísicos, geoquímicos e toda a espécie de dados produzida por *scanner* multiespectrais. Estes dados podem ser organizados em formatos não comprimidos (Figura 55). Conservar cada dado de uma variável como um arranjo em separado é um método conhecido como Band Sequencial (BSQ). Neste caso, um arquivo conteria o arranjo, por exemplo, a elevação, outro arquivo separado, o arranjo das classes de solo, outro, também separado, o mapa

geológico. Outra alternativa é chamada de Band Interleaved by Pixel (BIP), na qual todas as informações sobre uma pixel são conservadas juntas. A terceira alternativa é chamada de Band Interleaved by Line (BIL), na qual todos os valores de uma variável de uma simples linha são armazenados seguindo os valores das outras variáveis das linhas em sucessão.

Quando operações envolvem um simples tema ou um *layer* (Plano de Informação) de cada vez, o BSQ é o formato mais eficaz, porque o tema específico pode ser manipulado como uma entidade física independente. Quando, entretanto, necessita-se operar com mais de um tema ao mesmo tempo, o BIP é mais eficaz. Por exemplo, se tivermos as informações temáticas consistindo em elevação e classes de florestas e quisermos derivar a declividade, o BSQ deve ser usado, uma vez que essa informação pode ser produzida diretamente da elevação, sem a leitura do arquivo de classes de florestas. Se, por outro lado, quisermos comparar os dois temas, por exemplo, para encontrar as áreas onde as mais altas elevações estejam associadas às classes de florestas, o BIP é recomendado, porque os valores dos dois temas para cada pixel são adjacentes no banco de dados.

O BIL é um tipo de formatação intermediária entre os extremos BSQ e BIP e é usado em um grande número de sistemas raster comerciais. A conversão entre esses três tipos de formatação é relativamente simples.

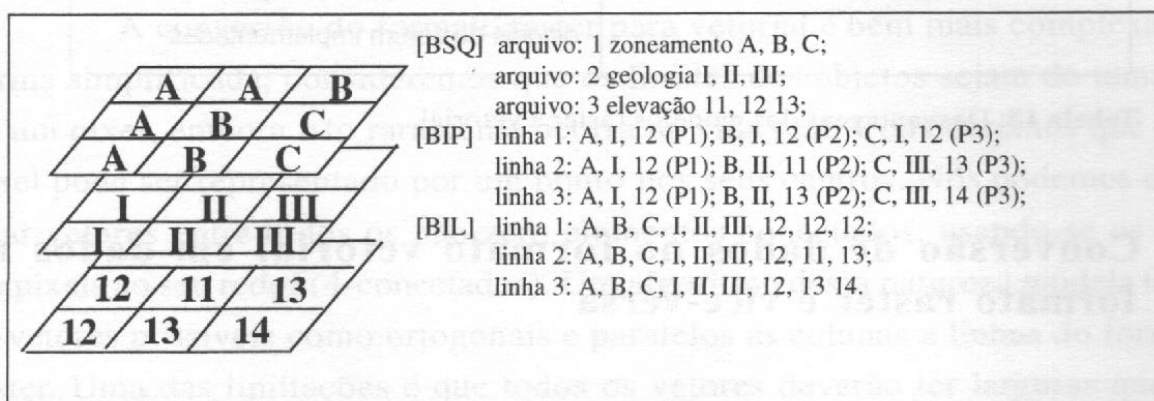


Figura 55: Representação de dados raster em formato BSQ, BIL e BIP.

Embora nos anos recentes tenha ocorrido um crescente aperfeiçoamento nos computadores, monitores e periféricos, os modelos raster e vetorial apre-



sentam características específicas e, naturalmente, vantagens e desvantagens, como é mostrado nas Tabelas 17 e 18.

MODELO RASTER	MODELO VETORIAL
Os dados possuem uma estrutura simples	Possui uma estrutura de dados compacta
Operações de superposição são facilmente implementadas	Permite uma codificação da topologia de forma eficaz, como resultado as análises de rede são facilmente implementadas
Altas variabilidades espaciais são eficientemente representadas	É recomendado para gráficos que devam se aproximar dos desenhos feitos à mão
Permite operações matemáticas com precisão	
Operações de modelagem e simulação são facilitadas	

**Tabela 17:** Vantagens dos modelos raster e vetorial.

MODELO RASTER	MODELO VETORIAL
A estrutura dos dados toma muito espaço de memória	A estruturação dos dados é complexa
As relações topológicas são difíceis de serem representadas	Operações de superposição são difíceis de serem implementadas
O produto final pode não ser esteticamente agradável	A representação de alta variabilidade espacial não é eficaz
	Operações de superposição são difíceis de serem implementadas

**Tabela 18:** Desvantagens dos modelos raster e vetorial.

#### 4.4 Conversão de dados no formato vetorial em dados no formato raster e vice-versa

Na grande maioria dos SIGs, há sempre a necessidade de realizar a conversão entre os formatos vetorial e raster para atender a objetivos específicos. Por exemplo, dados pontuais podem, via interpolação, ser transformados em superfícies contínuas, ou, por outro lado, dados escanerizados são transformados em dados vetoriais.

Existem numerosos problemas na conversão de dados vetoriais em raster e vice versa. A conversão de formato vetorial em raster é conceitualmente simples, embora praticamente difícil. No caso de pontos, o pixel correspondente no formato raster terá no seu centro as coordenadas geográficas e será codificado com o atributo do ponto. É pouco provável que a localização do ponto original coincida com o centro do pixel, dessa forma, pode-se perder a fenomenologia de determinados objetos. Este enfoque também ignora o problema de diferentes objetos ocupando o mesmo pixel. Por causa dessas limitações, a conversão do formato raster em vetorial não é usada freqüentemente.

No caso de elementos lineares, a mais simples estratégia de conversão consiste em identificar os pixels que são cortados pelas linhas e codificá-los com o atributo associado a essa linha. Com relação aos segmentos que não estão orientados com as linhas e colunas, a representação raster mostra uma distorção chamada **efeito escada**. Esta distorção pode ser minimizada utilizando-se dois procedimentos: com um sistema que possua uma maior resolução espacial ou modificando-se o contraste dos pixels adjacentes.

Os polígonos podem ser convertidos em formato raster em dois estágios. Primeiro os segmentos de linhas ou arcos que formam os limites dos polígonos são convertidos em raster tal qual foi descrito acima, produzindo o que chamamos de esqueleto do polígono. Segundo, os elementos raster contidos pelos limites do polígono são registrados com um atributo apropriado.

A conversão do formato raster para vetorial é bem mais complexa. De forma simplificada, consideremos que os limites dos objetos sejam do tamanho de um pixel, embora isto raramente ocorra na vida real. Consideremos que cada pixel pode ser representado por um ponto nos seus centros. Nós podemos desenhar vetores entre todos os não-zero elementos conectados, usando-se os quatro pixels ao seu redor (4-conectados). Um algoritmo dessa natureza modela todos os vetores possíveis como ortogonais e paralelos às colunas e linhas do formato raster. Uma das limitações é que todos os vetores deverão ter larguras que sejam um número inteiro e múltiplo da resolução espacial.

A transformação inversa não será possível com precisão, uma vez que a natureza da linha reta foi perdida no processo, não sendo mais possível recuperar o caráter da linha reta sem informações adicionais. Entretanto existem

certos algoritmos que podem extrair linhas retas de dados raster sob condições restritivas e com um custo adicional de computação.

Geralmente os SIGs em operação comercial têm duas funções gerais para a conversão: a esqueletização, seguida por uma estruturação matemática entre os objetos, que é determinada pelo reconhecimento dos nós e indicativos dos segmentos separados de todas as três figuras. Este algoritmo, entretanto, encontra muitas conexões entre os pixels que não são encontrados nos vetores originais; um melhoramento, envolvendo significativamente mais computação, seria somente desenhar linhas diagonais em quaisquer 3x3 pixels de vizinhos 8-conectados, quando elas estivessem ligadas por vetores ortogonais 4-conectados.

#### **4.5 Exemplos de dados em SIG**

Dados descrevendo o meio ambiente e relacionados aos recursos naturais e dados socioeconômicos podem ser analisados em um ambiente de SIG. No caso de dados relacionados aos recursos naturais, os SIGs podem ser utilizados como um repositório de um inventário para gerenciar, eficazmente, o potencial dos recursos, protegê-los contra atividades predatórias e modelar as complexas interações entre os fenômenos que possibilitam estabelecer previsões, que possam ser usadas como suporte para tomada de decisões.

Em regra, o conteúdo desses bancos de dados consiste em diferentes tipos de dados tais como geologia, vegetação, hidrologia, rede viária, limites políticos etc. Além destes, certamente os produtos derivados, como mapa de erodibilidade, produtividade do solo, fluxo de água subterrânea etc., devem ser parte componente do banco de dados.

Este banco de dados possui características específicas: é razoavelmente estático, portanto a atualização não é muito freqüente, a resolução espacial pode ser relativamente baixa, isto é, os pixels podem cobrir grandes áreas, a manipulação dos dados é feita utilizando-se de superfícies contínuas, portanto os diversos PIs correspondem a imagens no formato raster.

As fontes dos dados são derivadas de várias agências, por exemplo, geologia — Serviço Geológico Brasileiro; solos — EMBRAPA etc. Os mapas topo-

gráficos fornecem informações da altitude, redes viária e hídrica e os limites políticos. Outra fonte são os dados derivados de sensoriamento remoto por satélite como aqueles provenientes do Landsat TM, Radar, SPOT (Système Probatoire de Orbite Terrestre) etc.

Um exemplo de um banco de dados dessa natureza é o MLMIS (Minnesota Land Management Information) que consiste em um inventário envolvendo todo o estado e tem como objetivo o planejamento e o gerenciamento do meio físico. Possui uma listagem dos resultados na área de planejamento alcançados nos últimos 15 anos e a resolução espacial dos dados é de 100 m.

O banco de dados socioeconômico corresponde aos dados referentes às atividades humanas e espaço e/ou as estruturas utilizadas para conduzir atividades humanas. Este banco de dados inclui dados demográficos, disposição das unidades de residência ou comércio, dados de migração, transporte etc. As fontes dos dados são censos, sejam eles relacionados à população (educação, sexo, qualidade das casas etc.), econômicos (classes de indústria, tipos de comércio etc.), à agricultura (produção anual das colheitas, data de plantio etc.), aos registros das terras (proprietário das terras, descrição das propriedades etc.) ou ao inventário da infra-estrutura de transporte (ruas, avenidas, gasodutos, esgotos etc.).

Os produtos derivados deste banco de dados podem ser, entre outros, controle do recolhimento dos impostos, mapas específicos apresentando áreas de zoneamentos, registros geográficos atualizados das mudanças decorrentes da dinâmica do desenvolvimento socioeconômico, planejamento da saúde pública, da segurança pública e definição de áreas para fomento.

#### **4.6 Saída de dados**

O planejamento da saída gráfica dos resultados derivados da análise via SIG é uma tarefa que exige cuidados específicos, mesmo sabendo-se que a maioria dos produtos finais derivados da análise não será candidata à prateleira de bibliotecas e sim para ser usada imediatamente. A maioria dos SIGs é concebida para gerar produtos gráficos, nos quais as cores são selecionadas de maneira aleatória.

Não se deve esquecer que as cores utilizadas nos atlas cartográficos foram otimizadas a partir da experiência adquirida pelos cartógrafos ao longo de séculos. Um dos princípios básicos de excelência gráfica é apresentar ao observador o maior número de idéias, no menor tempo possível, com a menor quantidade de tinta, no menor espaço possível (Tufte, 1983).

As feições mostradas nos mapas podem ser exibidas e identificadas nos mapas de diferentes formas: símbolos, cores, tamanhos e textos. Na cartografia, o posicionamento do texto é um trabalho complexo e sofisticado. A tentativa de posicionar textos no monitor revela ao usuário uma grande dificuldade causada pela baixa resolução do monitor (p. ex.: 640 por 480 pixels).

Existem certos procedimentos padrões, como mostram as regras básicas de Imhof (1975), para gerar um produto cartográfico de boa qualidade, em que são seguidos os seguintes princípios: os nomes precisam ser legíveis, ser facilmente associados com as feições que descrevem, não se sobreporem a outras entidades no mapa, refletir a hierarquia das feições através de diferentes tamanhos de fontes, não produzir áreas com grande densidade de nomes, nem tampouco ser colocados de forma esparsa.

A descrição de feições pontuais deve ser colocada acima e à direita; abaixo e à direita é menos aceito; não é utilizada a colocação à esquerda. Descrições colocadas na vertical devem ser em pouca quantidade. A identificação de polígonos deve ser feita com as palavras centralizadas e, eventualmente, podem ser orientadas ou encurvadas para acompanhar as formas dos polígonos.

A escala dos arquivos de saída deve ser compatível e consoante as escalas dos arquivos de entrada. Assim, deve-se evitar digitalizar na escala de 1:1.000.000 e planejar a saída na escala de 1:25.000. Da mesma forma, não se deve digitalizar na escala 1:25.000 tendo em mente uma saída na escala de 1:1.000.000.

## Referências bibliográficas

- ARONOFF, S. (1989) *Geographic Information Systems: a Management Perspective*. WDL Publications, Ottawa, Canada.
- BURROUGH, P. A. (1986) *Principles of Geographical Information Systems: Methods and Requirements for Landuse Planning*. Clarendon, Oxford.
- IMHOF, E. (1975) "Positioning names on maps". *The American Cartographer* 2(2): 128-44.
- OEEPE (1984) *Test of digitising methods*. European Organisation for Experimental Photogrammetric Research. Official Publ. nº 14, Institut für Angewandte Geodäsie, Berlin.
- STAR, J. L. e ESTES, J. E. (1990) *Geographic Information Systems: An Introduction*, Prentice Hall.
- TUFTE, E. R. (1983) *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT. Contains numerous examples of graphical excellence (and its opposite) in map design.

## Bibliografia complementar

- FREEMAN, H. e AHN, J. (1984) "AUTOMAP — an expert system for automatic map name placement", *Proceedings, First International Symposium on Spatial Data Handling*, Zurich.
- ROBINSON, A. H., SALE, R. D., MORRISON, J. L. e MUEHRCKE, P. C. (1984) *Elements of Cartography*, 5th edition, Wiley, New York. Excellent source of map design principles.
- ZORASTER, S. (1986) "Integer programming applied to the map label placement problem", *Cartographica* 23(3): 16-27.